Wikipedia usage survey results

Issa Rice

2016-10-05

Contents

Summary

The summary is not intended to be comprehensive. It highlights the most important takeaways you should get from this post.

- Vipul Naik and I are interested in understanding how people use Wikipedia. One reason is that we are getting more people to work on editing and adding content to Wikipedia. We want to understand the impact of these edits, so that we can direct efforts more strategically. We are also curious!

- From May to July 2016, we conducted two surveys of people's Wikipedia usage. We collected survey responses from audience segments include Slate Star Codex readers, Vipul's Facebook friends, and a few audiences through SurveyMonkey Audience and Google Consumer Surveys. Our survey questions measured how heavily people use Wikipedia, what sort of pages they read or expected to find, the relation between their search habits and Wikipedia, and other actions they took within Wikipedia.

- Different audience segments responded very differently to the survey. Notably, the SurveyMonkey audience (which is closer to being representative of the general population) appears to use Wikipedia a lot less than Vipul's Facebook friends and Slate Star Codex readers. Their consumption of Wikipedia is also more passive: they are less likely to explicitly seek Wikipedia pages when searching for a topic, and less likely to engage in additional actions on Wikipedia pages. Even the college-educated SurveyMonkey audience used Wikipedia very little.

- This is tentative evidence that Wikipedia consumption is skewed towards a certain profile of people (and Vipul's Facebook friends and Slate Star Codex readers sample much more heavily from that profile). Even more tentatively, these heavy users tend to be more "elite" and influential. This tentatively led us to revise upward our estimates of the social value of a Wikipedia pageview.

- This was my first exercise in survey construction. I learned a number of lessons about survey design in the process.

- All the survey questions, as well as the breakdown of responses for each of the audience segments, are described in this post. Links to PDF exports of response summaries are at the end of the post.

Background

At the end of May 2016, Vipul Naik[1] and I created a Wikipedia usage survey[2] to gauge the usage habits of Wikipedia readers and editors. SurveyMonkey[3] allows the use of different "collectors" (i.e. survey URLs that keep results separate), so we circulated several different URLs among four locations to see how different audiences would respond. The audiences were as follows:

- SurveyMonkey's United States audience with no demographic filters (62 responses, 54 of which are full responses)
- Vipul Naik's timeline (post asking people to take the survey[4]; 70 responses, 69 of which are full responses). For background on Vipul's timeline audience, see his page on how he uses Facebook[5].
- The Wikipedia Analytics mailing list[6] (email linking to the survey[7]; 7 responses, 6 of which are full responses). Note that due to the small size of this group, the results below should not be trusted, unless possibly when the votes are decisive.
- Slate Star Codex (post that links to the survey[8]; 618 responses, 596 of which are full responses). While Slate Star Codex isn't the same as LessWrong, we think there is significant overlap in the two sites' audiences (see e.g. the recent LessWrong diaspora survey results[9]).
- In addition, although not an actual audience with a separate URL, several of the tables we present below will include an "H" group; this is the heavy users group of people who responded by saying they read 26 or more articles per week on Wikipedia. This group has 179 people: 164 from Slate Star Codex, 11 from Vipul's timeline, and 4 from the Analytics mailing list.

We ran the survey from May 30 to July 9, 2016 (although only the Slate Star Codex survey had a response past June 1).

After we looked at the survey responses on the first day, Vipul and I decided to create a second survey to focus on the parts from the first survey that interested

---

[1]http://lesswrong.com/user/VipulNaik/overview/
[2]https://meta.wikimedia.org/wiki/Research:Wikipedia_usage_survey
[3]https://en.wikipedia.org/wiki/SurveyMonkey
[4]https://www.facebook.com/vipulnaik.r/posts/10208540131276697
[5]http://vipulnaik.com/facebook/
[6]https://lists.wikimedia.org/mailman/listinfo/analytics
[7]https://lists.wikimedia.org/pipermail/analytics/2016-May/005219.html
[8]http://slatestarcodex.com/2016/06/02/links-616-linkandescence/
[9]http://lesswrong.com/lw/nor/2016_lesswrong_diaspora_survey_analysis_part_two/

us the most. The second survey was only circulated among SurveyMonkey's audiences: we used SurveyMonkey's US audience with no demographic filters (54 responses), as well as a US audience of ages 18–29 with a college or graduate degree (50 responses). We first ran the survey on the unfiltered audience again because the wording of our first question was changed and we wanted to have the new baseline. We then chose to filter for young college-educated people because our prediction was that more educated people would be more likely to read Wikipedia (the SurveyMonkey demographic data does not include education, and we hadn't seen the Pew Internet Research surveys in the next section, so we were relying on our intuition and some demographic data from past surveys) and because young people in our first survey gave more informative free-form responses in survey 2 (SurveyMonkey's demographic data *does* include age).

We ran a third survey on Google Consumer Surveys with a *single question* that was a word-to-word replica of the first question from the second survey. The main motivation here was that on Google Consumer Surveys, a single-question survey costs only 10 cents per response, so it was possible to get to a large number of responses at relatively low cost, and achieve more confidence in the tentative conclusions we had drawn from the SurveyMonkey surveys.

Previous surveys

Several demographic surveys regarding Wikipedia have been conducted, targeting both editors and users. The surveys we found most helpful were the following:

- The 2010 Wikipedia survey[10] by the Collaborative Creativity Group and the Wikimedia Foundation. The explanation before the bottom table on page 7 of the overview PDF[11] has "Contributors show slightly but significantly higher education levels than readers", which provides weak evidence that more educated people are more likely to engage with Wikipedia.
- The Global South User Survey 2014[12] by the Wikimedia Foundation
- Pew Internet Research's 2011 survey[13]: "Education level continues to be the strongest predictor of Wikipedia use. The collaborative encyclopedia is most popular among internet users with at least a college degree, 69% of whom use the site." (page 3)
- Pew Internet Research's 2007 survey[14]

---

[10]https://web.archive.org/web/20130717211630/http://wikipediastudy.org/

[11]https://web.archive.org/web/20131209060146/http://wikipediastudy.org/docs/Wikipedia_Overview_15March2010-FINAL.pdf

[12]https://upload.wikimedia.org/wikipedia/commons/8/8a/Global_South_User_Survey_2014_-_Full_Analysis_Report.pdf

[13]http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Wikipedia.pdf

[14]http://www.pewinternet.org/2007/04/24/wikipedia-users/

Note that we found the Pew Internet Research surveys after conducting our own two surveys (and during the write-up of this document).

Motivation

Vipul and I ultimately want to get a better sense of the value of a Wikipedia pageview (one way to measure the impact of content creation), and one way to do this is to understand how people are using Wikipedia. As we focus on getting more people to work on editing Wikipedia[15] – thus causing more people to read the content we pay and help to create – it becomes more important to understand what people are doing on the site.

For some previous discussion, see also Vipul's answers to the following Quora questions:

- What are the various parameters that affect the value of a pageview?[16]
- What's the relative social value of 1 Quora pageview (as measured by Quora stats http://www.quora.com/stats) and 1 Wikipedia pageview (as measured at, say, Wikipedia article traffic statistics)?[17]

Wikipedia allows relatively easy access to pageview data (especially by using tools developed for this purpose, including one that Vipul made[18]), and there are some surveys that provide demographic data (see "Previous surveys" above). However, after looking around, it was apparent that the kind of information our survey was designed to find was not available.

I should also note that we were driven by our curiosity of how people use Wikipedia.

Survey questions for the first survey

For reference, here are the survey questions for the first survey. A dummy/mock-up version of the survey can be found here: https://www.surveymonkey.com/r/PDTTBM8.

The survey introduction said the following:

> This survey is intended to gauge Wikipedia use habits. This survey has 3 pages with 5 questions total (3 on the first page, 1 on the second

---

[15] https://www.facebook.com/vipulnaik.r/posts/10208522775642817

[16] https://www.quora.com/What-are-the-various-parameters-that-affect-the-value-of-a-pageview/answer/Vipul-Naik

[17] https://www.quora.com/Whats-the-relative-social-value-of-1-Quora-pageview-as-measured-by-Quora-stats-http-www-quora-com-stats-and-1-Wikipedia-pageview-as-measured-at-say-Wikipedia-article-traffic-statistics/answer/Vipul-Naik

[18] http://wikipediaviews.org/

page, 1 on the third page). Please try your best to answer all of the questions, and make a guess if you're not sure.

And the actual questions:

1. How many distinct Wikipedia pages do you read per week on average?

   - less than 1
   - 1 to 10
   - 11 to 25
   - 26 or more

2. On a search engine (e.g. Google) results page, do you explicitly seek Wikipedia pages, or do you passively click on Wikipedia pages only if they show up at the top of the results?

   - I explicitly seek Wikipedia pages
   - I have a slight preference for Wikipedia pages
   - I just click on what is at the top of the results

3. Do you usually read a particular section of a page or the whole article?

   - Particular section
   - Whole page

4. How often do you do the following? (Choices: Several times per week, About once per week, About once per month, About once per several months, Never/almost never.)

   - Use the search functionality on Wikipedia
   - Be surprised that there is no Wikipedia page on a topic

5. For what fraction of pages you read do you do the following? (Choices: For every page, For most pages, For some pages, For very few pages, Never. These were displayed in a random order for each respondent, but displayed in alphabetical order here.)

   - Check (click or hover over) at least one citation to see where the information comes from on a page you are reading
   - Check how many pageviews a page is getting (on an external site or through the Pageview API)
   - Click through/look for at least one cited source to verify the information on a page you are reading
   - Edit a page you are reading because of grammatical/typographical errors on the page
   - Edit a page you are reading to add new information
   - Look at the "See also" section for additional articles to read
   - Look at the editing history of a page you are reading
   - Look at the editing history solely to see if a particular user wrote the page
   - Look at the talk page of a page you are reading

- Read a page mostly for the "Criticisms" or "Reception" (or similar) section, to understand different views on the subject
- Share the page with a friend/acquaintance/coworker

For the SurveyMonkey audience, there were also some demographic questions (age, gender, household income, US region, and device type).

Survey questions for the second survey

For reference, here are the survey questions for the second survey. A dummy/mock-up version of the survey can be found here: https://www.surveymonkey.com/r/28BW78V.

The survey introduction said the following:

> This survey is intended to gauge Wikipedia use habits. Please try your best to answer all of the questions, and make a guess if you're not sure.

> This survey has 4 questions across 3 pages.

> In this survey, "Wikipedia page" refers to a Wikipedia page in any language (not just the English Wikipedia).

And the actual questions:

1. How many distinct Wikipedia pages do you read (at least one sentence of) per week on average?

   - Fewer than 1
   - 1 to 10
   - 11 to 25
   - 26 or more

2. Which of these articles have you read (at least one sentence of) on Wikipedia (select all that apply)? (These were displayed in a random order except the last option for each respondent, but displayed in alphabetical order except the last option here.)

   - Adele
   - Barack Obama
   - Bernie Sanders
   - China
   - Donald Trump
   - Google
   - Hillary Clinton
   - India
   - Japan
   - Justin Bieber

- Justin Trudeau
- Katy Perry
- Taylor Swift
- The Beatles
- United States
- World War II
- None of the above

3. What are some of the Wikipedia articles you have most recently read (at least one sentence of)? Feel free to consult your browser's history.

4. Recall a time when you were surprised that a topic did not have a Wikipedia page. What were some of these topics?

Survey questions for the third survey (Google Consumer Surveys)

This survey had exactly one question. The wording of the question was exactly the same as that of the first question of the second survey.

1. How many distinct Wikipedia pages do you read (at least one sentence of) per week on average?

   - Fewer than 1
   - 1 to 10
   - 11 to 25
   - 26 or more

One slight difference was that whereas in the second survey, the order of the options was fixed, the third survey did a 50/50 split between that order and the exact reverse order. Such splitting is a best practice to deal with any order-related biases, while still preserving the logical order of the options. You can read more on the questionnaire design page of the Pew Research Center[19].

Results

In this section we present the highlights from each of the survey questions. If you prefer to dig into the data yourself, there are also some exported PDFs below provided by SurveyMonkey. Most of the inferences can be made using these PDFs, but there are some cases where additional filters are needed to deduce certain percentages.

We use the notation "SnQm" to mean "survey n question m".

---

[19]http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/

S1Q1: number of Wikipedia pages read per week

Here is a table that summarizes the data for Q1:

Table 1: How many distinct Wikipedia pages do you read per week on average? SM = SurveyMonkey audience, V = Vipul Naik's timeline, SSC = Slate Star Codex audience, AM = Wikipedia Analytics mailing list.

Response	SM	V	SSC	AM
less than 1	42%	1%	1%	0%
1 to 10	45%	40%	37%	29%
11 to 25	13%	43%	36%	14%
26 or more	0%	16%	27%	57%

Here are some highlights from the first question that aren't apparent from the table:

- Of the people who read fewer than 1 distinct Wikipedia page per week (26 people), 68% were female even though females were only 48% of the respondents. (Note that gender data is only available for the SurveyMonkey audience.)

- Filtering for high household income ($150k or more; 11 people) in the SurveyMonkey audience, only 2 read fewer than 1 page per week, although most (7) of the responses still fall in the "1 to 10" category.

The comments indicated that this question was flawed in several ways: we didn't specify which language Wikipedias count nor what it meant to "read" an article (the whole page, a section, or just a sentence?). One comment questioned the "low" ceiling of 26; in fact, I had initially made the cutoffs 1, 10, 100, 500, and 1000, but Vipul suggested the final cutoffs because he argued they would make it easier for people to answer (without having to look it up in their browser history). It turned out this modification was reasonable because the "26 or more" group was a minority.

S1Q2: affinity for Wikipedia in search results

We asked Q2, "On a search engine (e.g. Google) results page, do you explicitly seek Wikipedia pages, or do you passively click on Wikipedia pages only if they show up at the top of the results?", to see to what extent people preferred Wikipedia in search results. The main implication to this for people who do content creation on Wikipedia is that if people do explicitly seek Wikipedia pages (for whatever reason), it makes sense to give them more of what they want. On the other hand, if people don't prefer Wikipedia, it makes sense to

update in favor of diversifying one's content creation efforts while still keeping in mind that raw pageviews indicate that content will be read more if placed on Wikipedia (see for instance Brian Tomasik's experience[20], which is similar to my own, or gwern's page comparing Wikipedia with other wikis[21]).

The following table summarizes our results.

Table 2: On a search engine (e.g. Google) results page, do you explicitly seek Wikipedia pages, or do you passively click on Wikipedia pages only if they show up at the top of the results? SM = SurveyMonkey audience, V = Vipul Naik's timeline, SSC = Slate Star Codex audience, AM = Wikipedia Analytics mailing list, H = heavy users (26 or more articles per week) of Wikipedia.

Response	SM	V	SSC	AM	H
Explicitly seek Wikipedia	19%	60%	63%	57%	79%
Slight preference for Wikipedia	29%	39%	34%	43%	20%
Just click on top results	52%	1%	3%	0%	1%

One error on my part was that I didn't include an option for people who avoided Wikipedia or did something else. This became apparent from the comments. For this reason, the "Just click on top results" options might be inflated. In addition, some comments indicated a mixed strategy of preferring Wikipedia for general overviews while avoiding it for specific inquiries, so allowing multiple selections might have been better for this question.

S1Q3: section vs whole page

This question is relevant for Vipul and me because the work Vipul funds[22] is mainly whole-page creation. If people are mostly reading the introduction or a particular section like the "Criticisms" or "Reception" section (see S1Q5), then that forces us to consider spending more time on those sections, or to strengthen those sections on weak existing pages.

Responses to this question were fairly consistent across different audiences, as can be see in the following table.

---

[20]http://reducing-suffering.org/the-value-of-wikipedia-contributions-in-social-sciences/#Readership

[21]http://www.gwern.net/Wikipedia%20and%20Other%20Wikis

[22]https://github.com/vipulnaik/working-drafts/blob/master/new-article-pool.mediawiki

Table 3: Do you usually read a particular section of a page or the whole article? SM = SurveyMonkey audience, V = Vipul Naik's timeline, SSC = Slate Star Codex audience, AM = Wikipedia Analytics mailing list.

Response	SM	V	SSC	AM
Section	73%	80%	74%	86%
Whole	34%	23%	33%	29%

Note that people were allowed to select more than one option for this question. The comments indicate that several people do a combination, where they read the introductory portion of an article, then narrow down to the section of their interest.

S1Q4: search functionality on Wikipedia and surprise at lack of Wikipedia pages

We asked about whether people use the search functionality on Wikipedia because we wanted to know more about people's article discovery methods. The data is summarized in the following table.

Table 4: How often do you use the search functionality on Wikipedia? SM = SurveyMonkey audience, V = Vipul Naik's timeline, SSC = Slate Star Codex audience, AM = Wikipedia Analytics mailing list, H = heavy users (26 or more articles per week) of Wikipedia.

Response	SM	V	SSC	AM	H
Several times per week	8%	14%	32%	57%	55%
About once per week	19%	17%	21%	14%	15%
About once per month	15%	13%	14%	0%	3%
About once per several months	13%	12%	9%	14%	5%
Never/almost never	45%	43%	24%	14%	23%

Many people noted here that rather than using Wikipedia's search functionality, they use Google with "wiki" attached to their query, DuckDuckGo's "!w" expression, or some browser configuration to allow a quick search on Wikipedia.

To be more thorough about discovering people's content discovery methods, we should have asked about other methods as well. We did ask about the "See also" section in S1Q5.

Next, we asked how often people are surprised that there is no Wikipedia page

on a topic to gauge to what extent people notice a "gap" between how Wikipedia exists today and how it *could* exist. We were curious about what articles people specifically found missing, so we followed up with S2Q4.

> Table 5: How often are you surprised that there is no Wikipedia page on a topic? SM = SurveyMonkey audience, V = Vipul Naik's timeline, SSC = Slate Star Codex audience, AM = Wikipedia Analytics mailing list, H = heavy users (26 or more articles per week) of Wikipedia.

Response	SM	V	SSC	AM	H
Several times per week	2%	0%	2%	29%	6%
About once per week	8%	22%	18%	14%	34%
About once per month	18%	36%	34%	29%	31%
About once per several months	21%	22%	27%	0%	19%
Never/almost never	52%	20%	19%	29%	10%

Two comments on this question (out of 59) – both from the SSC group – specifically bemoaned deletionism, with one comment calling deletionism "a cancer killing Wikipedia".

S1Q5: behavior on pages

This question was intended to gauge how often people perform an action for a specific page; as such, the frequencies are expressed in page-relative terms.

The following table presents the scores for each response, which are weighted by the number of responses. The scores range from 1 (for every page) to 5 (never); in other words, the lower the number, the more frequently one does the thing.

> Table 6: For what fraction of pages you read do you do the following? Note that the responses have been shortened here; see the "Survey questions" section for the wording used in the survey. Responses are sorted by the values in the SSC column. SM = SurveyMonkey audience, V = Vipul Naik's timeline, SSC = Slate Star Codex audience, AM = Wikipedia Analytics mailing list, H = heavy users (26 or more articles per week) of Wikipedia.

Response	SM	V	SSC	AM	H
Check $\geq$ 1 citation	3.57	2.80	2.91	2.67	2.69
Look at "See also"	3.65	2.93	2.92	2.67	2.76
Read mostly for "Criticisms" or "Reception"	4.35	3.12	3.34	3.83	3.14
Click through $\geq$ 1 source to verify information	3.80	3.07	3.47	3.17	3.36

Response	SM	V	SSC	AM	H
Share the page	4.11	3.72	3.86	3.67	3.79
Look at the talk page	4.31	4.28	4.03	3.00	3.86
Look at the editing history	4.35	4.32	4.12	3.33	3.92
Edit a page for grammatical/typographical errors	4.50	4.41	4.22	3.67	4.02
Edit a page to add new information	4.61	4.55	4.49	3.83	4.34
Look at editing history to verify author	4.50	4.65	4.48	3.67	4.73
Check how many pageviews a page is getting	4.63	4.88	4.96	3.17	4.92

The table above provides a good ranking of how often people perform these actions on pages, but not the distribution information (which would require three dimensions to present fully). In general, the more common actions (scores of 2.5–4) had responses that clustered among "For some pages", "For very few pages", and "Never", while the less common actions (scores above 4) had responses that clustered mainly in "Never".

One comment (out of 43) – from the SSC group, but a different individual from the two in S1Q4 – bemoaned deletionism.

S2Q1: number of Wikipedia pages read per week

Note the wording changes on this question for the second survey: "less" was changed to "fewer", the clarification "at least one sentence of" was added, and we explicitly allowed any language. We have also presented the survey 1 results for the SurveyMonkey audience in the corresponding rows, but note that because of the change in wording, the correspondence isn't exact.

Table 7: How many distinct Wikipedia pages do you read (at least one sentence of) per week on average? SM = SurveyMonkey audience with no demographic filters, CEYP = College-educated young people of SurveyMonkey, S1SM = SurveyMonkey audience with no demographic filters from the first survey.

Response	SM	CEYP	S1SM
Fewer than 1	37%	32%	42%
1 to 10	48%	64%	45%
11 to 25	7%	2%	13%
26 or more	7%	2%	0%

Comparing SM with S1SM, we see that probably because of the wording, the percentages have drifted in the direction of more pages read. It might be surprising that the young educated audience seems to have a smaller fraction of heavy

users than the general population. However note that each group only had ~50 responses, and that we have no education information for the SM group.

S2Q2: multiple-choice of articles read

Our intention with this question was to see if people's stated or recalled article frequencies matched the actual, revealed popularity of the articles. Therefore we present the pageview data[23] along with the percentage of people who said they had read an article.

Table 8: Which of these articles have you read (at least one sentence of) on Wikipedia (select all that apply)? SM = SurveyMonkey audience with no demographic filters, CEYP = College-educated young people of SurveyMonkey. Columns "2016" and "2015" are desktop pageviews in millions. Note that the 2016 pageviews only include pageviews through the end of June. The rows are sorted by the values in the CEYP column followed by those in the SM column.

Response	SM	CEYP	2016	2015
None	37%	40%	—	—
World War II	17%	22%	2.6	6.5
Barack Obama	17%	20%	3.0	7.7
United States	17%	18%	4.3	9.6
Donald Trump	15%	18%	14.0	6.6
Taylor Swift	9%	18%	1.7	5.3
Bernie Sanders	17%	16%	4.3	3.8
Japan	11%	16%	1.6	3.7
Adele	6%	16%	2.0	4.0
Hillary Clinton	19%	14%	2.8	1.5
China	13%	14%	1.9	5.2
The Beatles	11%	14%	1.4	3.0
Katy Perry	9%	12%	0.8	2.4
Google	15%	10%	3.0	9.0
India	13%	10%	2.4	6.4
Justin Bieber	4%	8%	1.6	3.0
Justin Trudeau	9%	6%	1.1	3.0

Below are four plots of the data. Note that r_s denotes Spearman's rank correlation coefficient[24]. Spearman's rank correlation coefficient is used instead of Pearson's *r* because the former is less affected by outliers. Note also that the

---

[23]https://web.archive.org/web/20160714023739/http://wikipediaviews.org/displayviewsformultipleyears.php?tag=Pages%20in%20SurveyMonkey%20second%20survey
[24]https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

percentage of respondents who viewed a page counts each respondent once, whereas the number of pageviews does not have this restriction (i.e. duplicate pageviews count), so we wouldn't expect the relationship to be entirely linear even if the survey audiences were perfectly representative of the general population.
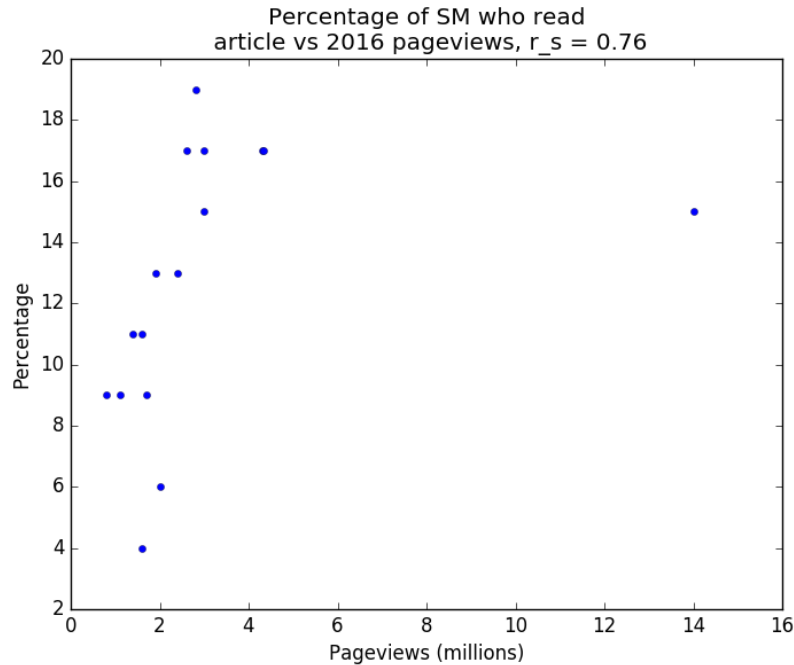

Figure 1: SM vs 2016 pageviews

S2Q3: free response of articles read

The most common response was along the lines of "None", "I don't know", "I don't remember", or similar. Among the more useful responses were:

- News stories (e.g. Death of Harambe[25], "WikiLeaks scandal" – unclear which page this is, since there are several pages on various aspects of WikiLeaks)
- Popular culture:
  - People including Megan Fox[26], LeBron James[27], Rita Hayworth[28]

---

[25]https://en.wikipedia.org/wiki/Death_of_Harambe
[26]https://en.wikipedia.org/wiki/Megan_Fox
[27]https://en.wikipedia.org/wiki/LeBron_James
[28]https://en.wikipedia.org/wiki/Rita_Hayworth

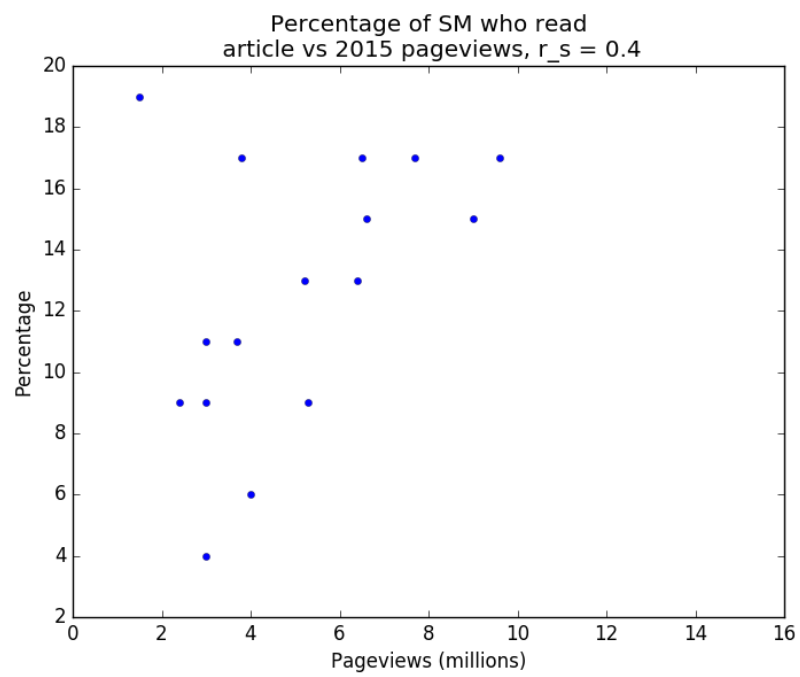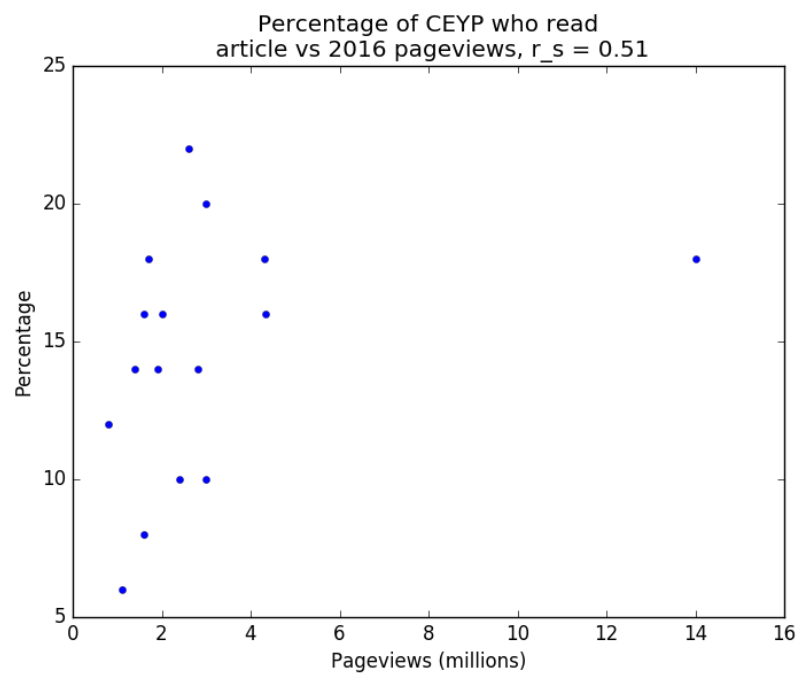Figure 2: SM vs 2015 pageviews
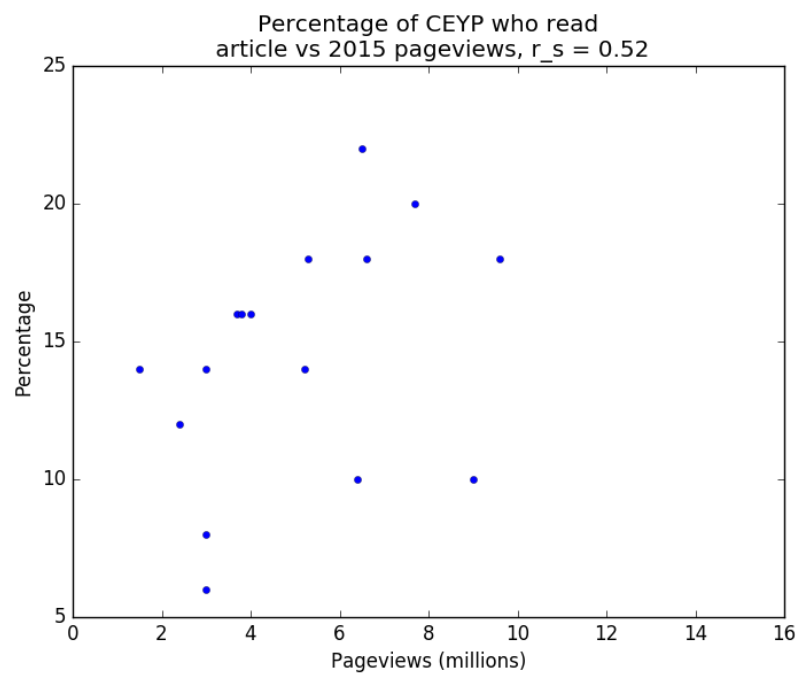
Figure 3: CEYP vs 2016 pageviews

Figure 4: CEYP vs 2015 pageviews

- – Works including Aladdin and the King of Thieves[29], X-Men: Apocalypse[30]
- More traditional encyclopedic information (e.g. Emerald ash borer[31], Spain[32], Siphonophorae[33], Scolopendra gigantea[34])

S2Q4: free response of surprise at lack of Wikipedia pages

As with the previous question, the most common response was along the lines of "None", "I don't know", "I don't remember", "Doesn't happen", or similar.

The most useful responses were classes of things: "particular words", "French plays/books", "Random people", "obscure people", "Specific list pages of movie genres", "Foreign actors", "various insect pages", and so forth.

S3Q1 (Google Consumer Surveys)

The survey was circulated to a target size of 500 in the United States (no demographic filters), and received 501 responses.

Since there was only one question, but we obtained data filtered by demographics in many different ways, we present this table with the columns denoting responses and the rows denoting the audience segments. We also include the S1Q1SM, S2Q1SM, and S2Q1CEYP responses for easy comparison. Note that S1Q1SM did not include the "at least one sentence of" caveat. We believe that adding this caveat would push people's estimates upward.

If you view the Google Consumer Surveys results online[35] you will also see the 95% confidence intervals for each of the segments. Note that percentages in a row may not add up to 100% due to rounding or due to people entering "Other" responses. For the entire GCS audience, every pair of options had a statistically significant difference, but for some subsegments, this was not true.

Audience segment	Fewer than 1	1 to 10	11 to 25	26 or more
S1Q1SM (N = 62)	42%	45%	13%	0%
S2Q1SM (N = 54)	37%	48%	7%	7%
S2Q1CEYP (N = 50)	32%	64%	2%	2%
GCS all (N = 501)	47%	35%	12%	6%

---

[29]https://en.wikipedia.org/wiki/Aladdin_and_the_King_of_Thieves
[30]https://en.wikipedia.org/wiki/X-Men:_Apocalypse
[31]https://en.wikipedia.org/wiki/Emerald_ash_borer
[32]https://en.wikipedia.org/wiki/Spain
[33]https://en.wikipedia.org/wiki/Siphonophorae
[34]https://en.wikipedia.org/wiki/Scolopendra_gigantea
[35]https://www.google.com/insights/consumersurveys/view?survey=o3iworx2rcfixmn2x5shtlppci&question=1&filter=&rw=1

Audience segment	Fewer than 1	1 to 10	11 to 25	26 or more
GCS male (N = 205)	41%	38%	16%	5%
GCS female (N = 208)	52%	34%	10%	5%
GCS 18–24 (N = 54)	33%	46%	13%	7%
GCS 25–34 (N = 71)	41%	37%	16%	7%
GCS 35–44 (N = 69)	51%	35%	10%	4%
GCS 45–54 (N = 77)	46%	40%	12%	3%
GCS 55–64 (N = 69)	57%	32%	7%	4%
GCS 65+ (N = 50)	52%	24%	18%	4%
GCS Urban (N = 176)	44%	35%	14%	7%
GCS Suburban (N = 224)	50%	34%	10%	6%
GCS Rural (N = 86)	44%	35%	14%	6%
GCS $0–24K (N = 49)	41%	37%	16%	6%
GCS $25–49K (N = 253)	53%	30%	10%	6%
GCS $50–74K (N = 132)	42%	39%	13%	6%
GCS $75–99K (N = 37)	43%	35%	11%	11%
GCS $100–149K (N = 11)	9%	64%	18%	9%
GCS $150K+ (N = 4)	25%	75%	0%	0%

We can see that the overall GCS data vindicates the broad conclusions we drew from SurveyMonkey data. Moreover, most GCS segments with a sufficiently large number of responses (50 or more) display a similar trend as the overall data. One exception is that younger audiences seem to be slightly less likely to use Wikipedia very little (i.e. fall in the "Fewer than 1" category), and older audiences seem slightly more likely to use Wikipedia very little.

Summaries of responses (exports for SurveyMonkey, weblink for Google Consumer Surveys)

SurveyMonkey allows exporting of response summaries. Here are the exports for each of the audiences.

- Survey 1, SurveyMonkey's audience[36]
- Survey 1, Vipul's timeline[37]
- Survey 1, Wikipedia Analytics mailing list[38]
- Survey 1, Slate Star Codex[39]
- Survey 1, Heavy users[40]

---

[36]http://files.issarice.com/SurveyMonkey.pdf

[37]http://files.issarice.com/Vipul_timeline.pdf

[38]http://files.issarice.com/Wikipedia_analytics_mailing_list.pdf

[39]http://files.issarice.com/Slate_Star_Codex.pdf

[40]http://files.issarice.com/Heavy_users.pdf

- Survey 2, no demographic filters[41]
- Survey 2, educated young people[42]

The Google Consumer Surveys survey results are available online at https://www.google.com/insights/consumersurveys/view?survey= o3iworx2rcfixmn2x5shtlppci&question=1&filter=&rw=1.

Survey-making lessons

Not having any experience designing surveys, and wanting some rough results quickly, I decided not to look into survey-making best practices beyond the feedback from Vipul. As the first survey progressed, it became clear that there were several deficiencies in that survey:

- Question 1 did not specify what counts as reading a page.
- We did not specify which language Wikipedias we were considering (multiple people noted how they read other language Wikipedias other than the English Wikipedia).
- Question 2 did not include an option for people who avoid Wikipedia or do something else entirely.
- We did not include an option to allow people to release their survey results.

Further questions

The two surveys we've done so far provide some insight into how people use Wikipedia, but we are still far from understanding the value of Wikipedia pageviews. Some remaining questions:

- Could it be possible that even on non-obscure topics, most of the views are by "elites" (i.e. those with outsized impact on the world)? This could mean pageviews are more valuable than previously thought.
- On S2Q1, why did our data show that CEYP was less engaged with Wikipedia than SM? Is this a limitation of the small number of responses or of SurveyMonkey's audiences?

---

[41]http://files.issarice.com/S2_unfiltered.pdf
[42]http://files.issarice.com/S2_educated.pdf

Further reading

- "The great decline in Wikipedia pageviews (condensed version)"[43] by Vipul Naik
- "In Defense Of Inclusionism"[44] by gwern

Acknowledgements

Thanks to Vipul Naik[45] for collaboration on this project and feedback while writing this document, and for supplying the summary section, and thanks to Ethan Bashkansky[46] for reviewing the document. All imperfections are my own.

The writing of this document was sponsored[47] by Vipul Naik. Vipul Naik also paid SurveyMonkey (for the cost of SurveyMonkey Audience) and Google Consumer Surveys.

Document source and versions

The source files used to compile this document are available in a GitHub Gist[48]. The Git repository of the Gist contains all versions of this document since its first publication.

This document is available in the following formats:

- As an HTML file at http://lesswrong.com/r/discussion/lw/nru/wikipedia_usage_survey_results/
- As a PDF file at http://files.issarice.com/wikipedia-survey-results.pdf

License

This document is released to the public domain[49].

---

[43]http://lesswrong.com/lw/lxc/the_great_decline_in_wikipedia_pageviews/

[44]http://www.gwern.net/In%20Defense%20Of%20Inclusionism

[45]http://vipulnaik.com/

[46]https://www.facebook.com/ethan.bashkansky

[47]https://github.com/vipulnaik/working-drafts/blob/master/contributor-lists/issa-list.mediawiki#Nonpage_payments

[48]https://gist.github.com/riceissa/7296b85303f96f2dc1dfb4f9a8ea44d9

[49]https://creativecommons.org/publicdomain/zero/1.0/